Alternate Accommodations Study of the AIMS Mathematics Exams

Prepared for the Arizona State Board of Education and Arizona Department of Education

By Jerome V. D'Agostino, Ph.D. & Deborah L. Rooks, M.S.Ed.

University of Arizona

Summary

Alternate Accommodations Study of the AIMS Mathematics Exams

A study was conducted to investigate if students' Arizona Instrument to Measure Standards (AIMS) mathematics scores were influenced by receiving acceptable alternate accommodations, which include the use of calculators. Data for the study were from the 2005 and 2006 spring AIMS mathematics exams for Grades Three through Eight and High School. Two separate but interrelated series of analyses were conducted. First, all operational items (including Dual Purpose Assessment Terra Nova items) on the exams from both years were studied for differential item functioning (DIF) related to alternate accommodation allowances. That is, the DIF analyses identified any item that was significantly easier or harder for students who were provided with alternate accommodations. Those items that were found to function differently were removed from the scoring process, and all students' revised scores were compared to their initial score based on the full set of items. The Winsteps computer program, which computes one-parameter item response theory (IRT) scores, was used for the analyses.

In the second series of analyses, all items that measured objectives considered by an expert committee convened by the Arizona Department of Education that would be influenced by calculator usage, were removed from the exams. As was conducted in the first phase of the study, students' initial scores were compared to their revised scores with those items removed.

DIF analysis requires one to define focal and reference groups from which item difficulties are computed. For this study, the focal students (Group 5) were those who received alternate accommodations according to Arizona guidelines, which included a four-function calculator, number chart, arithmetic table, manipulatives, or abacus. These alternate accommodations are available only to special education students whose Individual Education Programs specify such testing modifications. A challenge (and potential limitation) of this study

was to identify a reference group who was most comparable to the focal group in pertinent characteristics other than having a stipulated alternate accommodation. To combat this problem, four different reference groups were defined:

- Group 1: Students in general education without testing accommodations.
- Group 2: Students receiving special education services but without testing accommodations.
- Group 3: Students in general education receiving a standard testing accommodation.
- Group 4: Students receiving special education services and a standard testing accommodation.

As can be seen from the group identification criteria, Group 4 was perhaps the most similar to Group 5 as possible, given that Group 4 students were receiving special education services and were granted standard testing accommodations, which are defined by the Arizona Department of Education as "provisions made in how a student accesses and demonstrates learning," such as frequent breaks, extended time, large print, and use of an abacus. The groups were defined by grade and by test year. Tables 1a and 1b contain the average (and standard deviation) for each group per year, respectively. As can be seen from the tables, Group 1 was the most proficient group overall, while Group 5 was the least proficient. Though Group 4 was more proficient than the focal group, the differences were slight across the grades and two test years. Tables 2a and 2b provide the special education categorizations for Groups 4 and 5 for each year. As can be seen, the two groups had comparable special needs.

Besides calibrating item difficulties and person scores, Winsteps provides a built-in DIF analysis feature, which essentially compares item difficulties for focal and reference groups, after adjusting for the overall group proficiency levels. In a sense, DIF analysis addresses the

question, “Is Item X more or less difficult for one group versus another based on students who have comparable overall proficiency levels?” Items that demonstrated DIF either in favor of Group 5 (e.g., the item was easier for focal students) or against Group 5 (e.g., the item was harder for focal students) relative to any of the reference groups except Group 1 (general education students without accommodations) were flagged. A decision rule of 0.5 item calibration difference was used instead of statistical test results due to large sample sizes, as recommended in the Winsteps program manual.

Items that demonstrated DIF were removed from the scoring process, and revised student scores were produced without any DIF item included. Two revised scores were generated: one without items with DIF in favor of Group 5, and one without items with DIF against Group 5. Both revised scores were statistically compared to the initial students’ scores based on all operational items. This comparison was done for every student with a valid score regardless of group membership.

Tables 3a and 3b present the DIF results and impact on students’ scores by grade level and year. As can be seen from the tables, 2005 had slightly fewer DIF items than in 2006, but overall, the proportion of DIF items per test in most cases is under ten percent. The far right columns in both tables indicate the number of students whose scores were affected by removing DIF items. The tables reveal that not one student in either year was impacted either positively or negatively by the presence of DIF items. Typically, a test must contain a larger proportion of DIF items to affect students’ scores significantly.

In addition to the DIF analyses, results from the Arizona Department of Education’s study of calculator use impact were incorporated into the analyses. ADE asked a team of special education experts to review all mathematic objectives from the state standard and identify any

that might enhance or equalize a special education student's performance if provided a calculator. For this study, 2005 and 2006 items were located that measured the objectives selected by the committee as potentially enhancing a student's performance. Those items were removed from the test scoring process to compute another revised student score that was compared to the initial score. Further, the "enhanced" items also were examined for potential DIF to determine the convergence between the committee's decisions and the DIF analyses.

Tables 4a and 4b provide the results. Though the overall Arizona item pool contains a number of items representing objectives deemed by the committee as "enhanced," the 2005 and 2006 tests contained but a few of the items in the pool. Removing the items from the scoring process did not affect a single student's test score. Indeed, the great majority of the items measuring "enhanced" objectives did not demonstrate DIF.

The major limitations with this study were the difficulty in identifying a comparison group for students with alternate accommodations, and the limitation of extant data for identifying students who actually used a calculator. Thus, the results must be interpreted not as the impact of calculator use, but general alternate accommodation allowances. Though Group 4 seemed similar in terms of special education needs relative to Group 5, the former group was slightly more proficient overall compared to focal students.

Nonetheless, the results indicated very little influence of alternate accommodations. The few items that showed DIF had practically no impact on students' test performances, all but a few items from the pool that potentially could have enhanced students' performances actually were operational for the last two years, and the ones that were operational had little to no impact on scores. A decision to allow alternate mathematics accommodations as standard accommodations in Arizona would be congruent with the results of this study.

Table 1a

2005 Group means and standard deviations of math scale scores

	Group 1		*Group 2*		*Group 3*		*Group 4*		*Group 5*	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Grade 3	58691	452 (49)	3266	441 (53)	3180	417 (41)	2708	399 (40)	341	394 (39)
Grade 4	65297	482 (51)	3914	459 (55)	3268	437 (46)	3704	420 (42)	2179	418 (39)
Grade 5	66526	508 (53)	3523	478 (54)	3018	465 (42)	3817	443 (39)	2390	440 (34)
Grade 6	66485	522 (56)	3307	477 (53)	2803	477 (47)	3264	452 (40)	2405	447 (35)
Grade 7	68551	545 (52)	3180	497 (48)	2082	495 (48)	3104	479 (39)	2475	474 (34)
Grade 8	67848	558 (57)	3169	502 (50)	1867	510 (53)	2606	488 (44)	2709	478 (37)
High School	93746	700 (43)	3884	657 (31)	5340	676 (38)	2618	656 (31)	6152	654 (30)

Table 1b

2006 Group means and standard deviations of math scale scores

	Group 1		*Group 2*		*Group 3*		*Group 4*		*Group 5*	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Grade 3	68543	452 (47)	4313	435 (52)	3817	413 (38)	3800	399 (37)	713	395 (35)
Grade 4	68759	489 (52)	4142	465 (56)	3212	441 (43)	4622	422 (40)	909	421 (40)
Grade 5	68187	511 (52)	3919	480 (53)	2747	460 (40)	4852	446 (35)	1070	442 (31)
Grade 6	68956	524 (56)	3593	482 (54)	2711	471 (43)	4466	454 (38)	1280	446 (30)
Grade 7	69078	550 (54)	3200	499 (49)	2572	497 (45)	3688	481 (40)	1707	471 (32)
Grade 8	69327	561 (57)	3109	509 (51)	2473	507 (46)	3620	488 (42)	1841	481 (36)
High School	85230	698 (41)	4071	657 (29)	3262	674 (36)	4379	653 (26)	4844	652 (23)

Table 2a

Group 4 and Group 5 Comparisons

Special Education Categories	2005													
	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		High School	
	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5
Speech Impairment	41%	42%	36%	36%	30%	25%	24%	25%	20%	2%	14%	19%	8%	10%
Mild Mental Handicap	3%	4%	3%	5%	3%	4%	3%	4%	3%	4%	3%	5%	2%	4%
Learning Disability	73%	65%	76%	75%	79%	78%	78%	78%	80%	79%	80%	78%	79%	79%
Emotional Disability	8%	7%	8%	6%	8%	7%	9%	9%	11%	10%	12%	10%	11%	10%
Moderately Mental Handicap	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%

Table 2a continued

	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		High School	
Visual Impairment	1%	2%	1%	1%	1%	1%	1%	0%	1%	1%	1%	1%	1%	1%
Hearing Impairment	2%	4%	2%	2%	2%	1%	2%	1%	2%	1%	3%	1%	5%	1%
Other Health Impairment	6%	10%	7%	7%	7%	7%	7%	6%	7%	7%	6%	6%	6%	5%
Orthopedic Impairment	0%	0%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Traumatic Brain Injury	0%	1%	0%	0%	0%	0%	0%	1%	0%	1%	1%	1%	1%	1%
Multiple Disabilities	1%	1%	0%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Dual Sensory Impairment	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Autism	2%	4%	2%	3%	2%	2%	2%	2%	1%	2%	1%	1%	1%	1%
Severe Mental Handicap	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 2b

Group 4 and Group 5 Comparisons

Special Education Categories	2006													
	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		High School	
	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5	Group 4	Group 5
Speech Impairment	43%	44%	36%	41%	32%	37%	23%	28%	21%%	23%	18%	19%	1%	1%
Mild Mental Handicap	3%	5%	3%	6%	4%	6%	3%	5%	3%	4%	4%	4%	3%	3%
Learning Disability	68%	63%	74%	65%	75%	68%	76%	70%	75%	74%	73%	72%	78%	75%
Emotional Disability	7%	5%	7%	6%	7%	6%	9%	7%	10%	9%	12%	10%	10%	10%
Moderately Mental Handicap	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 2b continued

	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		High School	
Visual Impairment	1%	0%	1%	1%	1%	1%	1%	1%	1%	0%	1%	1%	1%	1%
Hearing Impairment	2%	2%	2%	2%	2%	2%	2%	1%	2%	1%	2%	1%	2%	1%
Other Health Impairment	7%	6%	7%	8%	8%	7%	8%	8%	8%	6%	8%	7%	6%	6%
Orthopedic Impairment	1%	2%	1%	2%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Traumatic Brain Injury	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	1%
Multiple Disabilities	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Dual Sensory Impairment	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Autism	3%	4%	2%	4%	2%	2%	2%	3%	2%	2%	2%	2%	1%	1%
Severe Mental Handicap	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 3a

Items Demonstrating DIF, 2005

		# of DIF items in favor of Group 5		# of DIF items against Group 5		Students Affected
	Total # of Items		% of total		% of total	
Grade 3	72	0	0%	9	12.50%	0
Grade 4	70	2	0%	0	0%	0
Grade 5	78	4	5.13%	0	0%	0
Grade 6	68	1	1.47%	0	0%	0
Grade 7	68	2	2.94%	0	0%	0
Grade 8	66	5	7.58%	0	0%	0
High School	85	0	0%	2	2.35%	0

Table 3b

Items Demonstrating DIF, 2006

		# of DIF items in favor of Group 5		# of DIF items against Group 5		Students Affected
	Total # of Items		% of total		% of total	
Grade 3	72	6	8.33%	4	5.56%	0
Grade 4	70	7	10%	0	0%	0
Grade 5	68	6	8.82%	2	2.94%	0
Grade 6	68	0	0%	2	2.94%	0
Grade 7	68	2	2.94%	0	0%	0
Grade 8	66	5	7.58%	0	0%	0
High School	84	0	0%	0	0%	0

Table 4a

2005 Calculator Enhanced Items

Mathematics Standard	# of total items on assessment	# items in pool		# of calculator sensitive items found on assessment		# of calculator sensitive items showing DIF	
	#	#	%	#	%	#	%
Grade 3	72						
1-2-4		2	2.78%	0	0%	0	0
Grade 4	70						
1-2-1		6	8.57%	1	1.43%	0	0
1-2-2		7	10%	1	1.43%	0	0
1-2-5		5	7.14%	1	1.43%	0	0
1-2-6		5	7.14%	1	1.43%	0	0
1-2-7		2	2.86%	0	0	0	0
1-2-13		4	5.71%	0	0	0	0
1-3-2		2	2.86%	1	1.43%	0	0
Grade 5	78						
1-2-3		4	10.84%	0	0	0	0
1-2-4		4	10.84%	0	0	0	0
1-2-12		4	10.84%	1	1.28%	1	1.28%
1-2-13		4	10.84%	0	0	0	0
1-2-14		3	3.85%	1	1.28%	0	0

Table 4a continued

2005 Calculator Enhanced Items

Mathematics Standard	# of total items on assessment	# items in pool		# of calculator sensitive items found on assessment		# of calculator sensitive items showing DIF	
	#	#	%	#	%	#	%
Grade 6	68						
1-2-4		2	2.94%	0	0	0	0
1-2-9		6	8.82%	1	1.47%	0	0
1-2-12		4	5.88%	1	1.47%	0	0
1-3-2		3	4.41%	0	0	0	0
1-3-5		3	4.41%	0	0	0	0
Grade 7	68						
1-2-1		2	2.94%	0	0	0	0
1-2-2		4	5.88%	1	1.47%	1	1.47%
1-2-5		5	7.35%	0	0	0	0
1-2-6		4	5.88%	0	0	0	0
1-2-10		6	8.82%	0	0	0	0
1-2-12		3	4.41%	0	0	0	0
1-3-2		1	1.47%	0	0	0	0
Grade 8	66						
1-1-1		5	7.58%	1	1.52%	0	0
1-2-4		1	1.52%	0	0	0	0
1-3-2		2	3.03%	0	0	0	0

Table 4a continued

2005 Calculator Enhanced Items

Mathematics Standard	# of total items on assessment	# items in pool		# of calculator sensitive items found on assessment		# of calculator sensitive items showing DIF	
	#	#	%	#	%	#	%
High School	85						
1-2-3		12	14.12%	1	1.14%	0	0
1-2-6		9	10.23%	0	0	0	0
1-2-7		7	7.95%	1	1.14%	0	0
1-3-2		3	3.41%	1	1.14%	0	0
1-3-3		7	7.95%	0	0	0	0

Table 4b

2006 Calculator Enhanced Items

Mathematics Standard	# of total items on assessment	# items in pool		# of calculator sensitive items found on assessment		# of calculator sensitive items showing DIF	
	#	#	%	#	%	#	%
Grade 3	72						
1-2-4		2	2.78%	0	0%	0	0
Grade 4	70						
1-2-1		6	8.57%	0	0	0	0
1-2-2		7	10%	1	1.43%	1	1.43%
1-2-5		5	7.14%	1	1.43%	0	0
1-2-6		5	7.14%	1	1.43%	0	0
1-2-7		2	2.86%	0	0	0	0
1-2-13		4	5.71%	0	0	0	0
1-3-2		2	2.86%	1	1.43%	0	0
Grade 5	68						
1-2-3		4	5.88%	0	0	0	0
1-2-4		4	5.88%	0	0	0	0
1-2-12		4	5.88%	1	1.47%	1	1.47%
1-2-13		4	5.88%	0	0	0	0
1-2-14		3	4.41%	1	1.47%	1	1.47%

Table 4b continued

2006 Calculator Enhanced Items

Mathematics Standard	# of total items on assessment	# items in pool		# of calculator sensitive items found on assessment		# of calculator sensitive items showing DIF	
	#	#	%	#	%	#	%
Grade 6	68						
1-2-4		2	2.94%	0	0	0	0
1-2-9		6	8.82%	1	1.47%	0	0
1-2-12		4	5.88%	1	1.47%	0	0
1-3-2		3	4.41%	0	0	0	0
1-3-5		3	4.41%	0	0	0	0
Grade 7	68						
1-2-1		2	2.94%	0	0	0	0
1-2-2		4	5.88%	0	0	0	0
1-2-5		5	7.35%	0	0	0	0
1-2-6		4	5.88%	0	0	0	0
1-2-10		6	8.82%	0	0	0	0
1-2-12		3	4.41%	0	0	0	0
1-3-2		1	1.47%	0	0	0	0
Grade 8	66						
1-1-1		5	7.58%	1	1.52%	0	0
1-2-4		1	1.52%	0	0	0	0
1-3-2		2	3.03%	0	0	0	0

Table 4b continued

2006 Calculator Enhanced Items

Mathematics Standard	# of total items on assessment	# items in pool		# of calculator sensitive items found on assessment		# of calculator sensitive items showing DIF	
	#	#	%	#	%	#	%
High School	84						
1-2-3		12	14.12%	0	0	0	0
1-2-6		9	10.23%	0	0	0	0
1-2-7		7	7.95%	0	0	0	0
1-3-2		3	3.41%	1	1.14%	0	0
1-3-3		7	7.95%	1	1.14%	0	0